

Weaving a Web of Data

SOASIST Chapter Meeting
Sept 12, 2007

Eric Miller
em@zepheira.com



me



dad





scientist





PI



Simile



(previously)



semantic web lead





director





Dublin Core Metadata Initiative®

Making it easier to find information.



(now)



president





about zepheira

- firm offering Semantic technologies solutions bridging technology, business and people
- team of leaders in next generation Web technologies and business applications
- founded in 2007, already featured in MIT Tech Review, BusinessWeek, eWeek, Investors Business Daily, etc.



common theme



decentralized data management



free data from
applications



reduce costs to find,
reuse, share information



accelerate business,
science and personal
discovery



my goals today



challenge



stimulate



show



make you curious



make you think



Pop Quiz



When did we get to Web
1.0 ?



the Web



pages and links





Southern Ohio Chapter

←

→

↺

ⓧ

🏠

http://www.asis.org/Chapters/soasis/meeting_9-12-07.html

🔍

🔍

Getting Started Latest Headlines

asis&t

The Information Society for the Information Age

sign in

contact us

site index

About ASIS&T

Membership

Conferences

Publications

SIGS & Chapters

Careers

E-Mail Lists

Go

Home > SIGS & Chapters > Chapters - Regional > Southern Ohio Chapter > SOH Meeting 9/12/07

Special Interest Groups (SIGs)

Chapters - Regional

Chapters - Student

Southern Ohio Chapter Meeting 9/12/07

The Southern Ohio Chapter of the American Society for Information Science & Technology

Presentation: Eric Miller on the Semantic Web

WHEN: Wednesday, September 12, 2007; 3:00 - 5:00 PM

WHERE: LexisNexis, 9443 Springboro Pike, Miamisburg, OH, Building 4, Auditorium 1

COST: \$25.00 nonmembers, \$15.00 ASIST and SLA members, \$5.00 students. Please pay at sign-in. Checks or cash accepted.

Known for his vision and leadership roles in the Dublin Core Metadata Initiative at OCLC and the Semantic Web Initiative for W3C at MIT, Eric Miller will discuss how leveraging Semantic Web standards can enable organizations achieve data interoperability for seamless integration of online products and often disparate product components in a Service Oriented Architecture environment. Eric will elucidate the RDF data model and the use of Microformats in surfacing online data.

Eric Miller is the President of Zepheira, The Art of Data, which provides solutions to effectively integrate, navigate and manage information across boundaries of person, group and enterprise. Dr. Miller is highly sought after as an adviser to businesses and other organizations and speaker at conferences worldwide, providing insights on the evolution of the Web.

Eric holds a Research Scientist position at MIT's Computer Science and Artificial Intelligence Laboratory where he is a Principal Investigator on the MIT SIMILE project focused on developing robust, open source tools based on Semantic Web technologies that improve access, management and reuse among digital resources.

Officers

[Arne Almqvist](#), Chair
[Jill Sellers](#), Secretary
[Rosemary Ashton](#), Treasurer
[Donald Case](#), Chapter Assembly Rep
[Rosemary Ashton](#), Alternate Chapter Rep.
[Jill Sellers](#), Past Chair

Contact Information

Web Site
<http://www.soasis.org/>

Find: 🔍

Next Previous Highlight all

Done

Southern Ohio Chapter

http://www.asis.org/Chapters/soasis/meeting_9-12-07.html

Getting Started Latest Headlines

asis&t
The Information Society for the Information Age

sign in contact us site index

About ASIS&T Membership Conferences Publications **SIGS & Chapters** Careers E-Mail Lists

Go

Home > SIGS & Chapters > Chapters - Regional > Southern Ohio Chapter Meeting 9/12/07

Special Interest Groups (SIGs)
Chapters - Regional
Chapters - Student

Organization

Southern Ohio Chapter Meeting 9/12/07

The Southern Ohio Chapter of the American Society for Information Science & Technology

Presentation: Eric Miller on the Semantic Web

WHEN: Wednesday, September 12, 2007; 3:00 - 5:00 PM
WHERE: LexisNexis, 9443 Springboro Pike, Miamisburg, OH, Building 4, Auditorium 1
COST: \$25.00 nonmembers, \$15.00 ASIST and SLA members, \$5.00 students. Please pay at sign-in. Checks or cash accepted.

Known for his vision and leadership roles in the Dublin Core Metadata Initiative at OCLC and the Semantic Web Initiative for W3C at MIT, Eric Miller will discuss how leveraging Semantic Web standards can enable organizations achieve data interoperability for seamless integration of online products and often disparate product components in a Service Oriented Architecture environment. Eric will elucidate the RDF data model and the use of Microformats in surfacing online data.

Eric Miller is the President of Zepheira, The Art of Data, which provides solutions to effectively integrate, navigate and manage information across boundaries of person, group and enterprise. Dr. Miller is highly sought after as an adviser to businesses and other organizations and speaker at conferences worldwide, providing insights on the evolution of the Web.

Eric holds a Research Scientist position at MIT's Computer Science and Artificial Intelligence Laboratory where he is a Principal Investigator on the MIT SIMILE project focused on developing robust, open source tools based on Semantic Web technologies that improve access, management and reuse among digital resources.

Officers

[Arne Almqvist](#), Chair
[Jill Sellers](#), Secretary
[Rosemary Ashton](#), Treasurer
[Donald Case](#), Chapter Assembly Rep
[Rosemary Ashton](#), Alternate Chapter Rep.
[Jill Sellers](#), Past Chair

Contact Information

Web Site
<http://www.soasis.org>

Event

People

Find: Next Previous Highlight all

Done

i want my data!



most of the web is

- designed for direct human consumption
- homogenous linking (“points at”)
- data is hidden



web trends

- people
- interfaces
- 're-useful' data



data rich, access poor

- no common format for representing data
- no easy way of polling site / query page



brute force



very few sites are done
by aggregating other web
site's data



for example





Google Search

I'm Feeling Lucky

[Advanced Search](#)
[Preferences](#)
[Language Tools](#)

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2007 Google

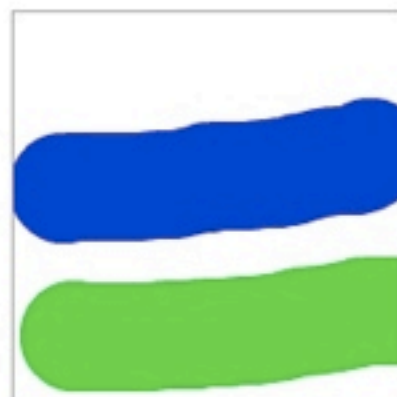
Welcome in 2006!

retrievr

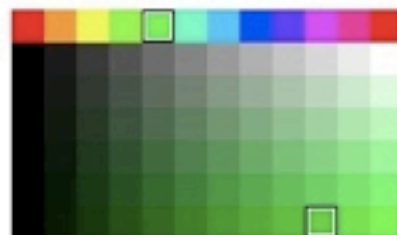
All images

Search by:

Sketch • [Image](#)



◀ Undo Redo ▶ Clear



New! Please help us [rate sketches](#) for The Art of retrievr! Only a million to go ...

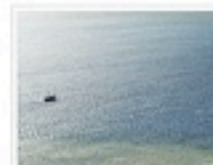
Still new! You can search by



From [T Glow](#)



From [T Glow](#)



From [Renata Diem](#)



From [phitar](#)



From [MARIO MARCANTE](#)



From [eyecatcher](#)



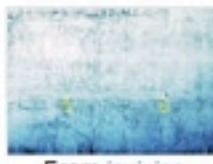
From [angelocesare](#)



From [Photos Marcante@](#)



From [Ahmed Zahid](#)



From [jayjuice](#)



From [h.neumeyer](#)



From [zomerstorm](#)



From [Thiago Lopes](#)



From [Charles Machado](#)



From [phil.h](#)



From [Katemina](#)



[Search](#) [Tags](#) [Blog Finder](#) [Explore](#)

Enron's Skilling and Lay convicted... [See what bloggers are saying](#)

Search the Blogosphere

SEARCH

[Options](#)

Explore the Blogosphere

[Explore All >>](#)

Tech

Guild Wars Vs WoW Vs MC Hammer
A Fan-made Music Video Sets Out to
Settle If Guild Wars or World of Warcraf...

Free Blogger Templates

If You Are Blogger/Blogspot Blogger, You
Can Download Some FREE High Qualit...

[Tech >](#)

Politics

Jefferson's History of Sleaze
I Never Realized What a Sleazy
Character William Jefferson Is -- I've
Been ...

JOHN LOGAN AND MEMORIAL DAY
5/28/2006 THE COURAGE OF ONE'S
CONVICTIONS AND WHERE IT'S LACKING
CATEGORY:....

[Politics >](#)

See also: [Popular: YouTube, News, Movies, Books](#) and [Top 100 blogs](#).

Sponsored link: [Virtuoso Net Solutions](#) Award Winning Web Hosting. Editor's Choice Awards '05, '06. [Your Ad Here](#)

Michelle Madhok's Favorite Blogs

Latest posts:

- Fire Magic Monarch
Magnum
- A Find: Diane Von
Furstenberg Eyelet
- Roberto Cavalli
Freedom Bag, Handbag

[View Favorites >](#)



What's happening right now

Currently tracking 41.7 million sites and 2.5 billion links.

Top Searches

- [Clay Aiken](#)
- [Swift Boating The Planet](#)
- [Lost](#)
- [Haditha](#)
- [Macbook](#)
- [Memorial Day](#)
- [Vallevershwan](#)

Hot Tags

- [memorial day](#)
- [Iraq](#)
- [web-20](#)
- [wordpress](#)
- [Sex](#)
- [iPod](#)
- [clayaiken](#)
- [Politics](#)

Featured Blog [Add Yours!](#)

[Randy's Publicity Newsgroup](#)



By Randy Mast.
Tagged Actors,
Authors, Blog,
Business, fashion, Journalism,
Marketing...

[More in Blog Finder >](#)

[Start search over](#)

Change your search

Departure airport:

CMH (Columbus) ▼

Destination airport:

BOS (Boston) ▼

Travel Dates:

Specific | [Flexible](#)

Departing: (mm/dd/yy)

5/29/2007

Anytime ▼

Returning: (mm/dd/yy)

5/30/2007

Anytime ▼

Airline: [More Info](#)

No Preference ▼

Class

Economy / Coach ▼

☐ Nonstop flights only

☐ Refundable flights only

only

Go

Change Travelers

1 Adult

[Change travelers](#)

✈ - Indicates flight is operated by another airline. Move your mouse over the icon for details.

Columbus, OH (CMH) to Boston, MA (BOS)

	All Results	US Airways	jetBlue Airways	nwa Northwest	Continental	AA American Airlines
Nonstop	from \$223 see below	---	from \$223	---	---	---
1 stop	from \$204 see below	from \$204	---	from \$244	from \$254	from \$264
2+ stops	---	---	---	---	---	---

[Show more airlines »](#)

Are your travel dates flexible? [See our Fare Comparison Calendar](#)

Note: The prices shown below are for the **flight only**; they are e-ticket prices and include [all flight taxes and fees](#). If your itinerary requires paper tickets there will be an [additional charge](#). These results cover a metro area with [several airports](#). Review your choices carefully.

jetBlue
AIRWAYS

JetBlue is now available for booking on Expedia!

Book JetBlue on Expedia and **save \$36** per roundtrip flight.

Hurry—offer ends May 30! [See details](#)

1 Choose a departing flight or [view complete roundtrips](#)

Sort by: ☒ Price ☐ Duration ☐ Departure time ☐ Arrival time

from **\$204 Roundtrip**

7:05 pm Depart Columbus (CMH)
Arrive Boston (BOS) **10:24 pm**

2 tickets left at this price! [See details](#)

Tue 29-May
Duration: 3hr 19mn



US Airways ✈ 3092 / 2050
Connect in Washington DC (DCA)

[➔ Choose this departure](#)

from **\$204 Roundtrip**

2 tickets left at this price! [See details](#)

Fare Alert
Get the best prices delivered right to your desktop.
[Sign up now](#)

Live Search

Image Search Made Easy!

Caribbean beaches



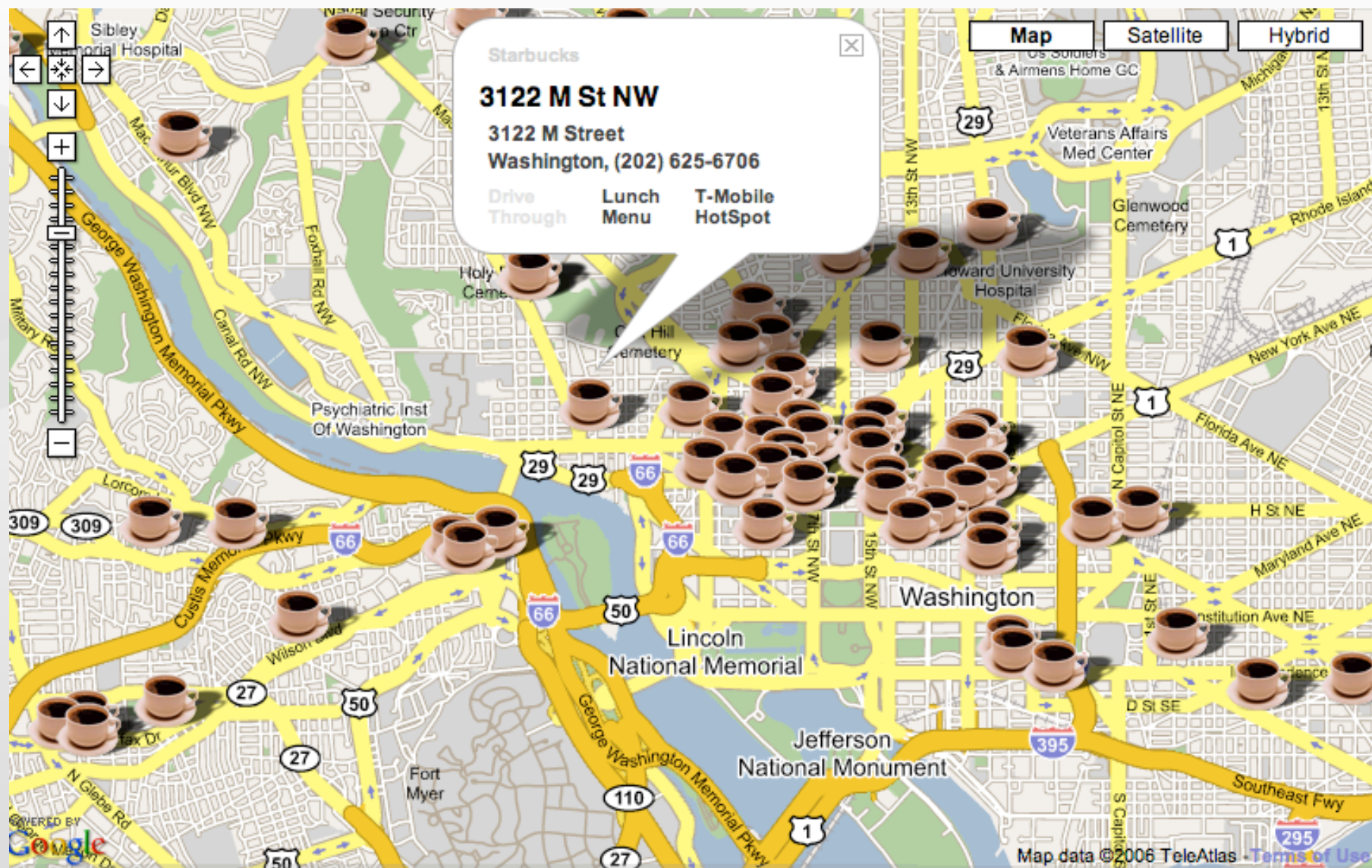
DRAG ME



mash-ups

- mix data from different sites to provide added value
- the mashed sites don't need to be involved
- hybrid client-server model - client on one side, server on the other



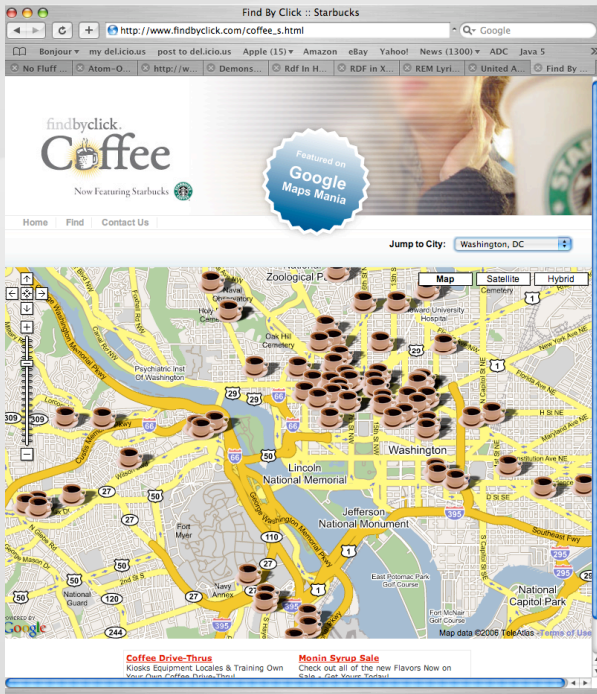


Data Dies in the Browser



Data Dies in the Browser

In Loving Memory
of Our Mashup
2/28/2007-2/28/2007



mash-up problems

- data is mostly locked up in pages
- each web site is different
- and keeps changing



mash-up problems

- very blurry lines between use and fair-use
- even after extraction, data needs to be modeled so that it can be mashed
- a mashed-up web site looks like another web site (so difficult for further mashing)



mashing in short

- extremely useful
- hard
- doesn't cascade



so what do we do?



web of data

- making mash-ups easier
- making cascades possible



semantic web



semantic web

- free data from application boundaries
- enabling data integration and reuse
- across application, departmental, organization and community boundaries
- capability to draw new conclusions
- share with others



example: where to stay?



Local Hotels and Motels

http://www.stanford.edu/dept/hds/chs/general/hotel.html

Getting Started Latest Headlines World Wide Web C... Danny Ayers The W3C Team Page Questions to the AC MIT Libraries RDF/A

Eric's Meetings Local Hotels and Motels

Community Housing Services
A division of Residential & Dining Enterprises
STANFORD UNIVERSITY

Contact Us | Search | FAQs | CHS | R&DE | SU home

Local Hotels and Motels

Search Rental Listings Post a Rental Listing Info for Tenants Info for Landlords About the Stanford Area Hotel List Off-campus subsidized apts

Please refer to the bottom of this table for a [key to codes and price ranges](#).

Miles from Campus	Address	Telephone	Price Range	Codes
Stanford-Owned-and-Operated				
2.2	SLAC Guest House , 2575 Sand Hill Road, Menlo Park	(650) 926-2800	\$ - \$\$	ST
Menlo Park				
1.4	Best Western Riviera , 15 El Camino, 94025	(650) 321-8772	\$\$\$	PKB
1.4	Stanford Park Hotel , 100 El Camino, 94025	(650) 322-1234	\$\$\$\$	PSTW
1.9	Mermaid Inn Motel , 727 El Camino, 94025	(650) 323-9481	\$	PKWB
2.2	SLAC Guest House , 2575 Sand Hill Road, 94025	(650) 926-2800	\$ - \$\$	ST
2.4	Menlo Park Inn, 1315 El Camino, 94025	(650) 326-7530	\$\$	KB
2.8	The Red Cottage , 1704 El Camino, 94025	(650) 326-9010	\$\$	PKSB
Palo Alto				
0.8	Sheraton Palo Alto Hotel , 625 El Camino, 94301	(650) 328-2800	\$\$\$\$	PKST
0.8	Westin Palo Alto, 675 El Camino, 94301	(650) 321-4422	\$\$\$\$	PST
1.0	The Cardinal Hotel , 235 Hamilton Avenue, 94301	(650) 323-5101	\$\$	KS
1.2	Stanford Terrace Inn , 531 Stanford Avenue, 94306	(650) 857-0333	\$\$\$	PKSBT
1.4	Garden Court Hotel , 520 Cowper Street, 94301	(650) 322-9000	\$\$\$\$	SWT
1.7	Garden Court Hotel , 520 Cowper Street, 94301	(650) 322-9000	\$\$\$\$	SWT

Done

Local Hotels and Motels

[http://www.stanford.edu/dept/hds/chs/general/hotel.htm](#)

[Getting Started](#)
[Latest Headlines](#)
[World Wide Web C...](#)
[Danny Ayers](#)
[The W3C Team Page](#)
[Questions to the AC](#)
[MIT Libraries](#)
[RDF/A](#)

Housing Services

A division of Residential & Dining Enterprises
STANFORD UNIVERSITY

Local Hotels and Motels

[Search Rental Listings](#)
[Post a Rental Listing](#)
[Info for Tenants](#)
[Info for Landlords](#)
[About the Stanford Area](#)
[Hotel List](#)
[Off-campus subsidized apts](#)

Please refer to the bottom of this table for a [key to codes and price ranges](#).

Miles from Campus	Address	Telephone	Price Range	Codes
Stanford-Owned-and-Operated				
2.2	SLAC Guest House, 2575 Sand Hill Road, Menlo Park	(650) 926-2800	\$ - \$\$	ST
Menlo Park				
1.4	Best Western Riviera, 15 El Camino, 94025	(650) 321-8772	\$\$\$	PKB
1.4	Stanford Park Hotel, 100 El Camino, 94025	(650) 322-1234	\$\$\$\$	PSTW
1.9	Mermaid Inn Motel, 727 El Camino, 94025	(650) 323-9481	\$	PKWB
		(650)		

Find:

[Find Next](#)
[Find Previous](#)
[Highlight all](#)
[Match case](#)

URLs

Code: 1

Insert

Run

```

var prefixRDF = 'http://www.w3.org/1999/02/22-r
var prefixDC = 'http://purl.org/dc/elements/1.1
var prefixLoc = 'http://simile.mit.edu/2005/05/

var namespace = doc.documentElement.namespaceURI
var nsResolver = namespace ? function(prefix) {
  if (prefix == 'x') return namespace; else ret
} : null;

var getNode = function(doc, contextNode, xpath,
  return doc.evaluate(xpath, contextNode, nsRes

```

Capture Console Results Help

/html/body/div/table/tbody/tr[2]/td/table/tbody/tr/td/ti

Item

Variable

Item 1 (FONT)

Item 2 (FONT)

http://www.stanford.edu/dept/hds/...

SLAC Guest Hou...

2575 Sand Hill Road, Menlo Park

Addr

Item 3 (FONT)

Item 4 (FONT)

Item 5 (FONT)

Item 6 (FONT)

Done

file:///Users/em/Library/Application%20Support/Fir... Collected Information

1 filter criterion

• type: Hotel (remove)

Order Commands

List View Calendar View Map View Graph View

33 items
sorted by address [A to Z]

1(3) 2(4) 3(3) 4(6) 5(2) 6(2) 7(2) c(2) h(1) m(2) p(1) q(1) s(1) t(2)

« previous 1 2 3 4 next »

Creekside Inn [URI]

address	3400 El Camino, CA, 94306
coordinates	37.419463,-122.135429
title	Creekside Inn
type	Hotel

Show Referers

Save Publish Tags: Tag

Rickey's Hyatt [URI]

address	4219 El, CA,
coordinates	37.0625,-95.677068
title	Rickey's Hyatt
type	Hotel

Find: Find Next Find Previous Highlight all Match case

Done

Type here to search

- type
- address
- coordinates (map)
- title



file:///Users/em/Library/Application%20Support/Fir...

Collected Information

1 filter criterion

- type: Hotel (remove)

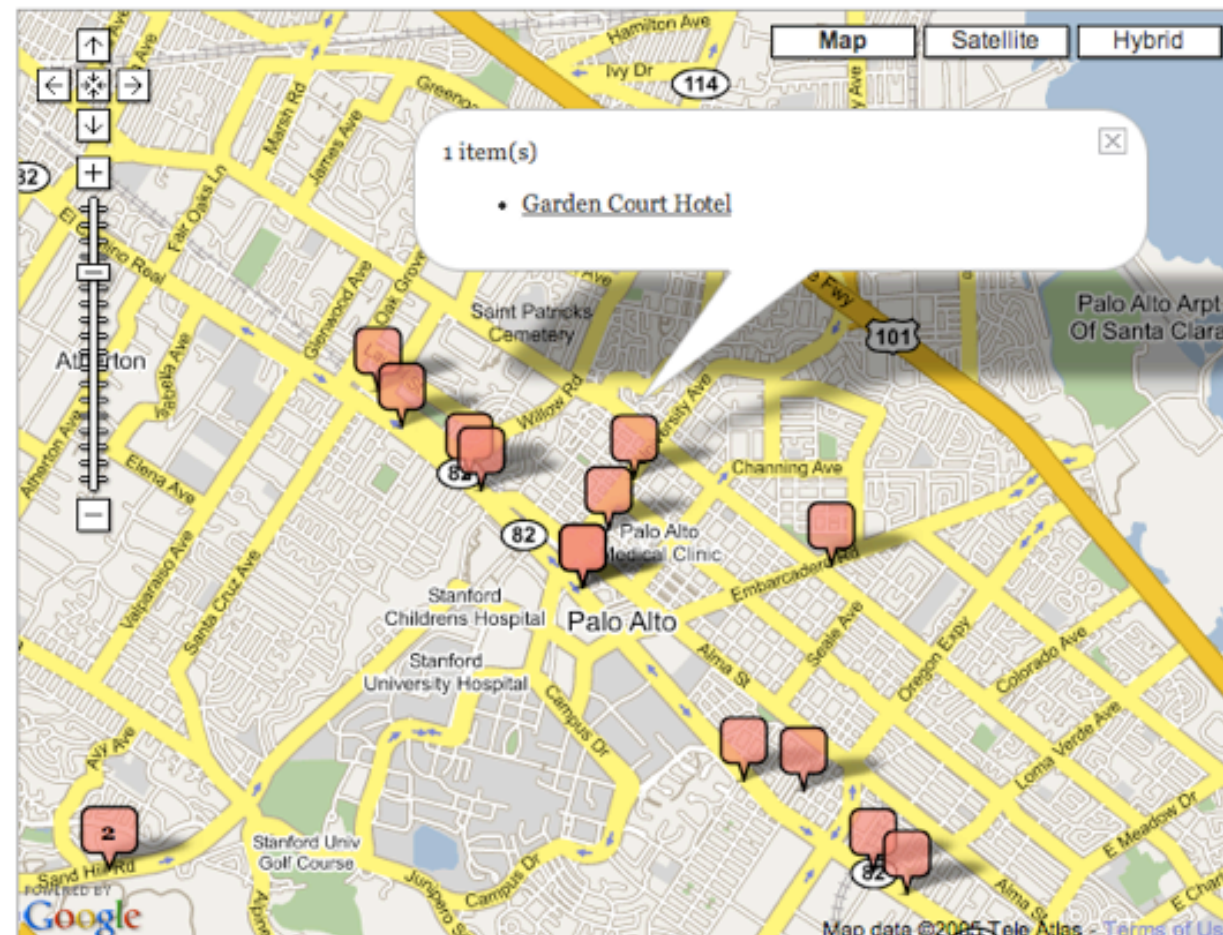
Order Commands

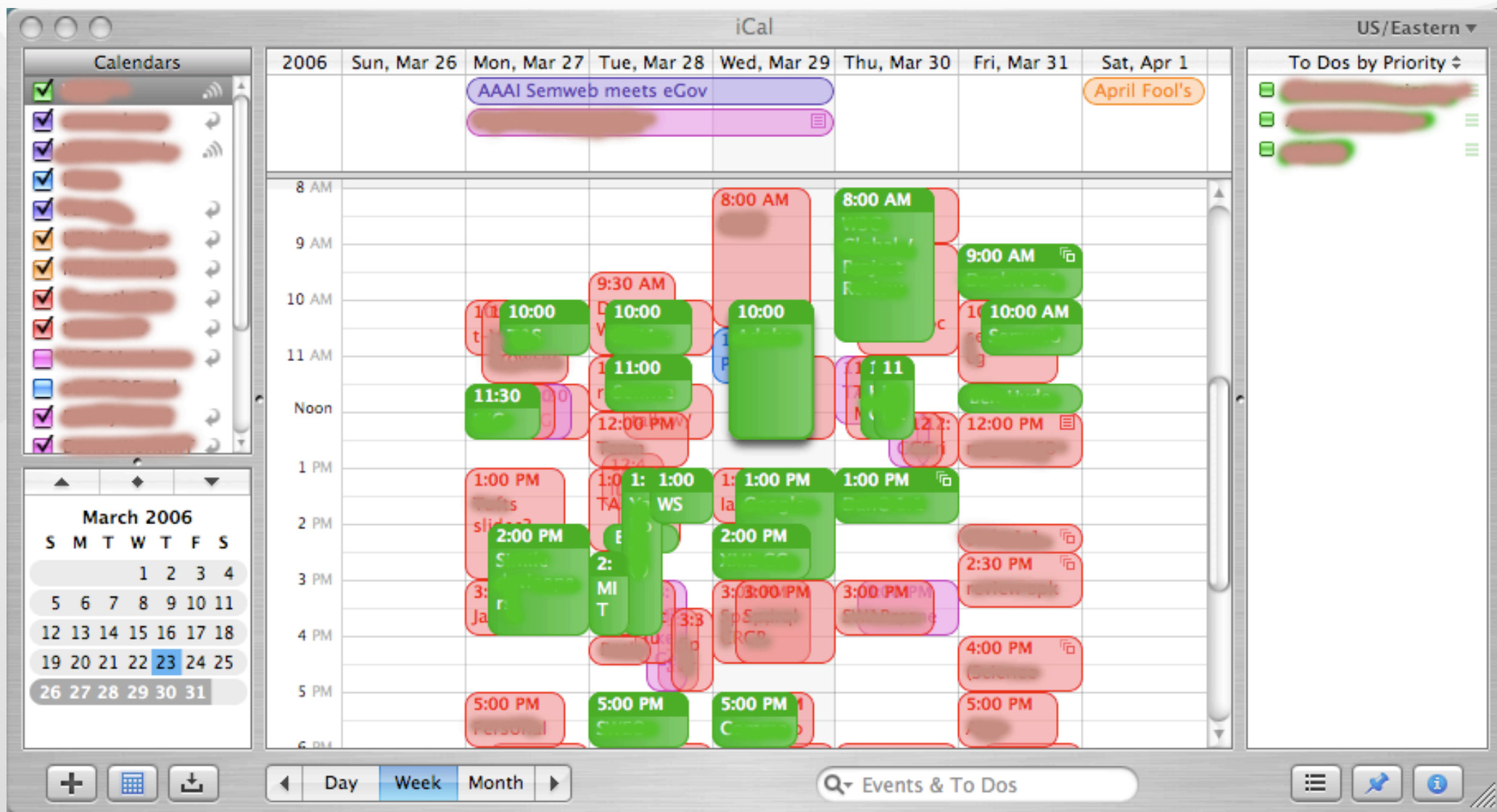
List View Calendar View Map View Graph View

Type here to search

- type
- address
- coordinates (map)
- title

Simile





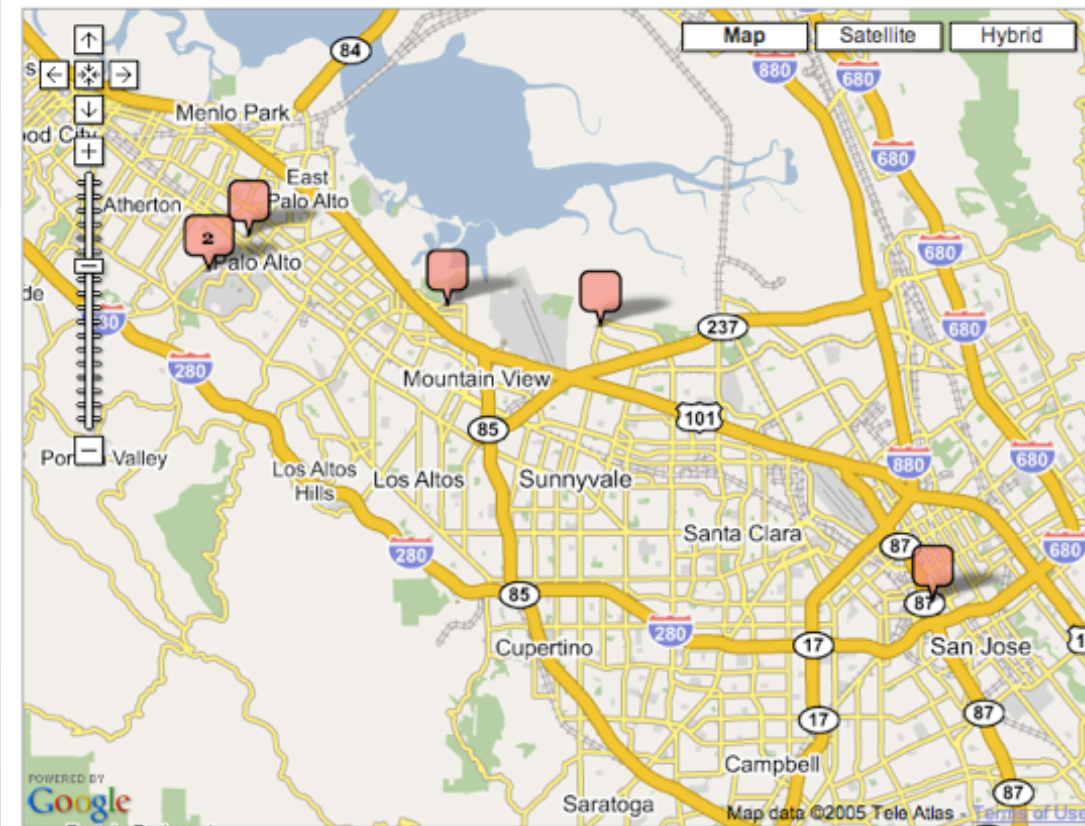
file:///Users/em/Library/Application%20Support/Fir... Collected Information

1 filter criterion

- type: Organization (remove)

Order Commands

List View Calendar View Map View Graph View



Organization

Find: Find Next Find Previous Highlight all Match case

Done

My Piggy Bank

←

→

↺

✖

🏠

📖

🐷

http://127.0.0.1:1978/piggy-bank/5af23b1c024cf383

🔍

Getting Started Latest Headlines World Wide Web C... Danny Ayers The W3C Team Page Questions to the AC MIT Libraries RDF/A

My Piggy Bank

My Piggy Bank

Mar 23, 2006 10:36:22 AM

1 filter criterion

- Tag: "hotels" (remove) "meetings" (remove) [add more]

Order Commands

List View Calendar View Map View Graph View

1 item(s)

- Garden Court Hotel

Map Satellite Hybrid

type

Type here to filter

Meetings (6)

📍 type

📍 tag

📍 coordinates (map)

📍 Title

📍 address

Simile

📍 Hotel

📍 Meetings

Find: 🔍

Find Next

Find Previous

Highlight all

Match case

Done

How is this done?



RDF



Resource Description Framework (RDF)

Subject	Predicate	Object
<u>http://mytalk.com</u>	dc:creator	“eric miller”

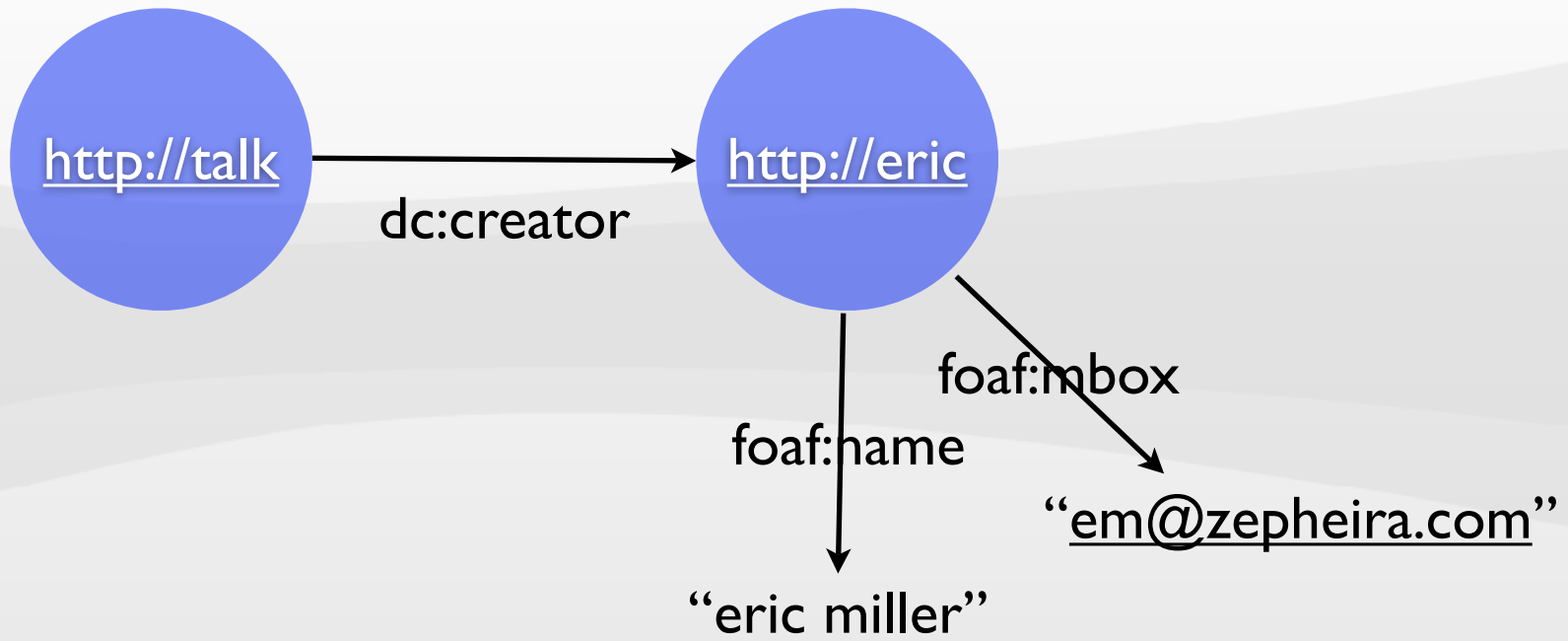
URI

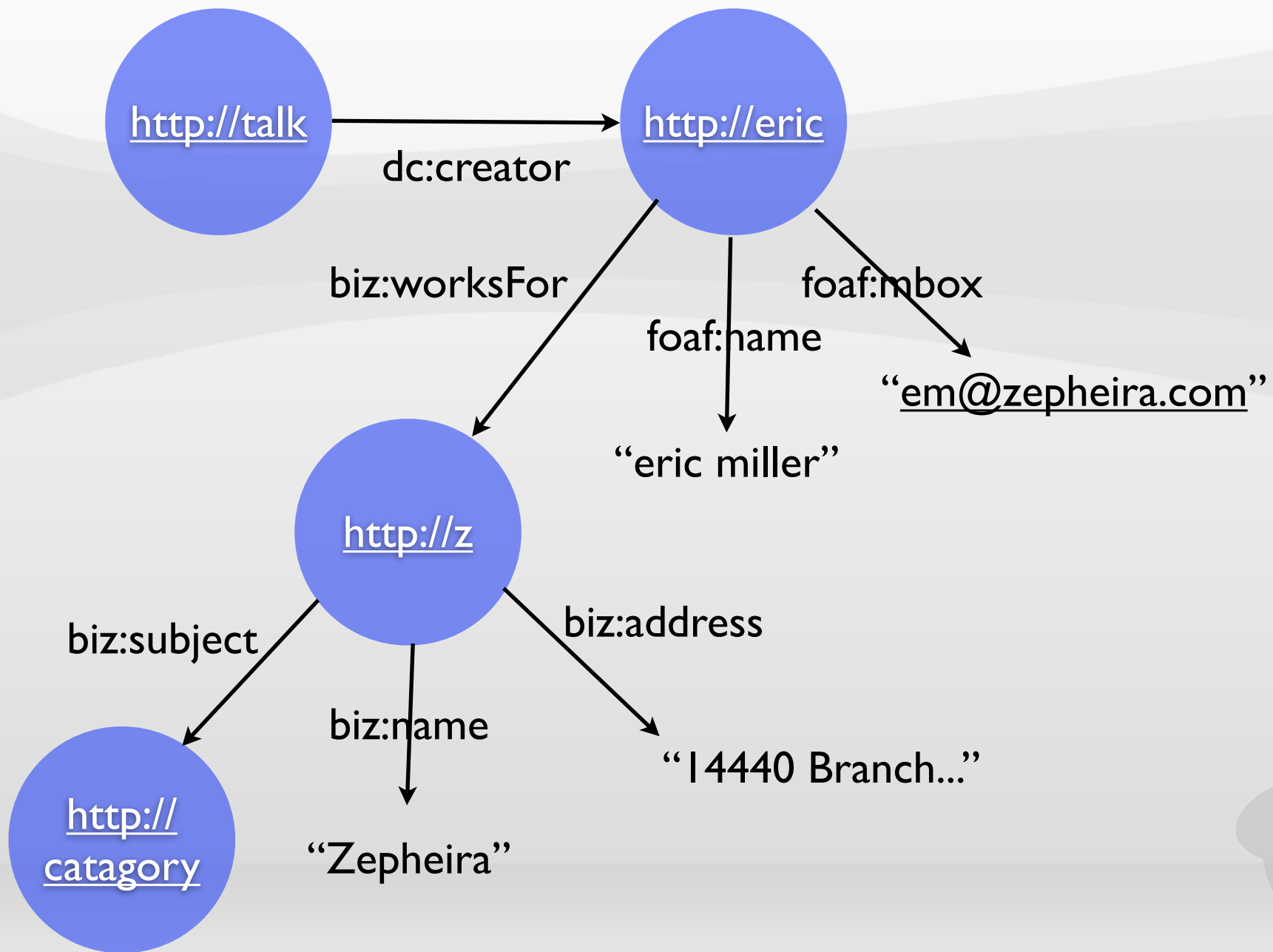
URI

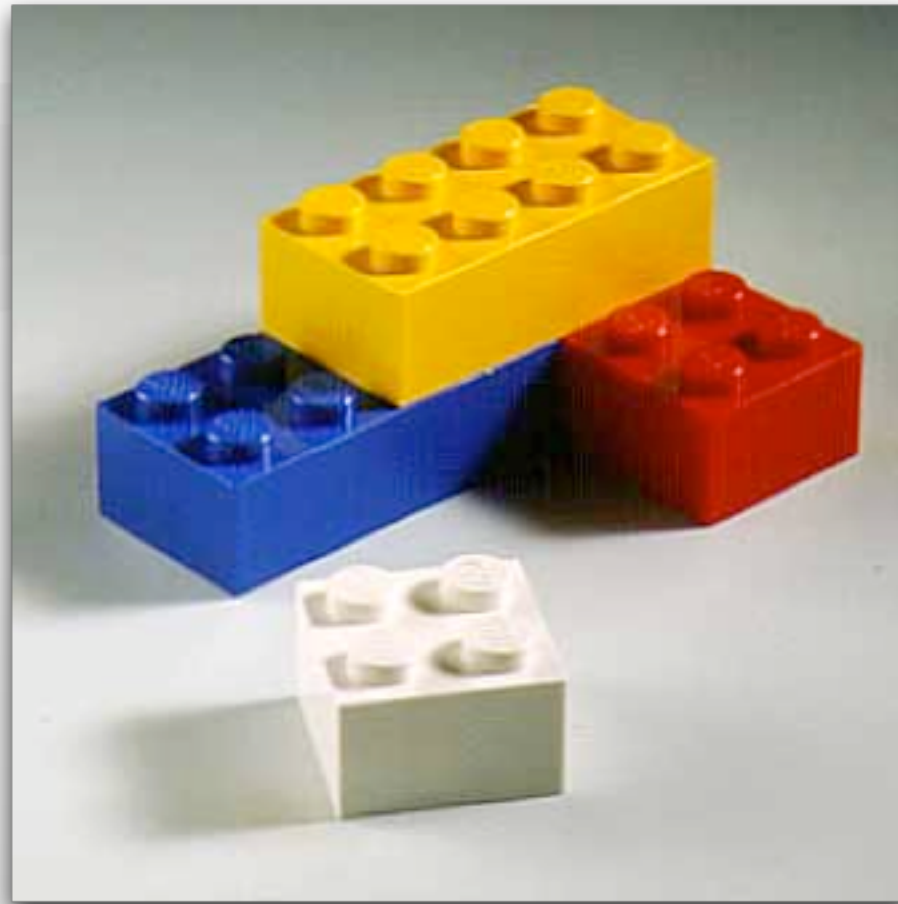
Literal





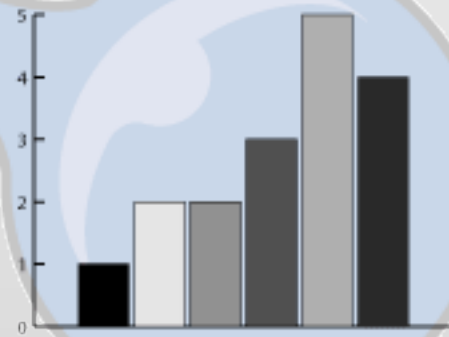








<XHTML>
<XML>



enabling 'data' technologies

- RDF
- OWL
- SKOS
- GRDDL
- SPARQL



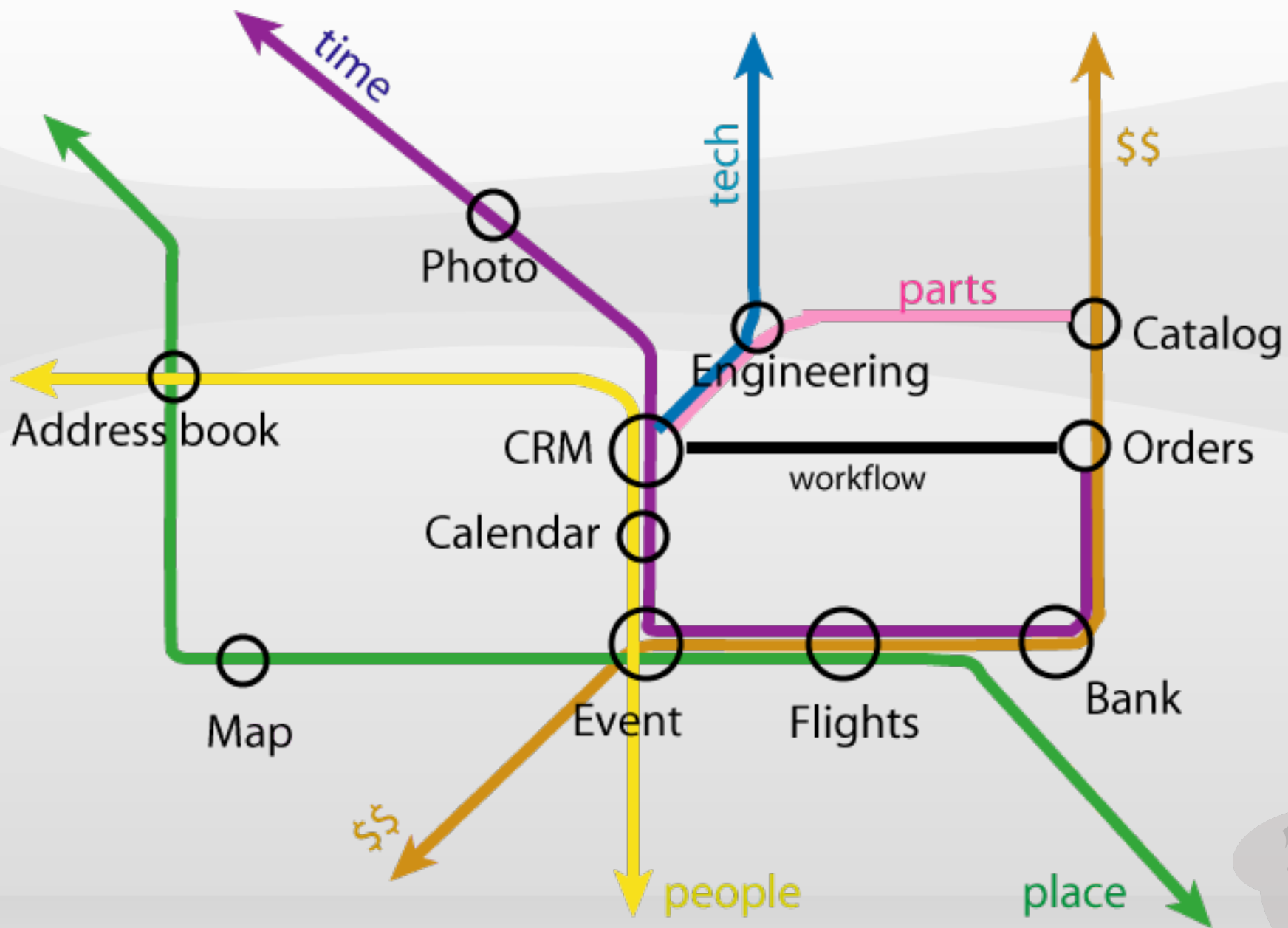
evolution and revolution

- Web
 - evolution not revolution
 - easier way of “joining” web sites
- Business
 - a quiet revolution
 - and about to get very loud



in the enterprise





recombinant data

- the ability to rapidly recombine, reform, re-factor and reuse data from different applications to address a particular task, need or objective
- across private, semi-private, public data sources



market indicators



oracle



adobe



nokia



eli lilly



garlik



joost



and many others...



strategic acquisitions



VC



even more open source
tools / stories



and coupled with new
rich client front ends
makes for some very
exciting possibilities



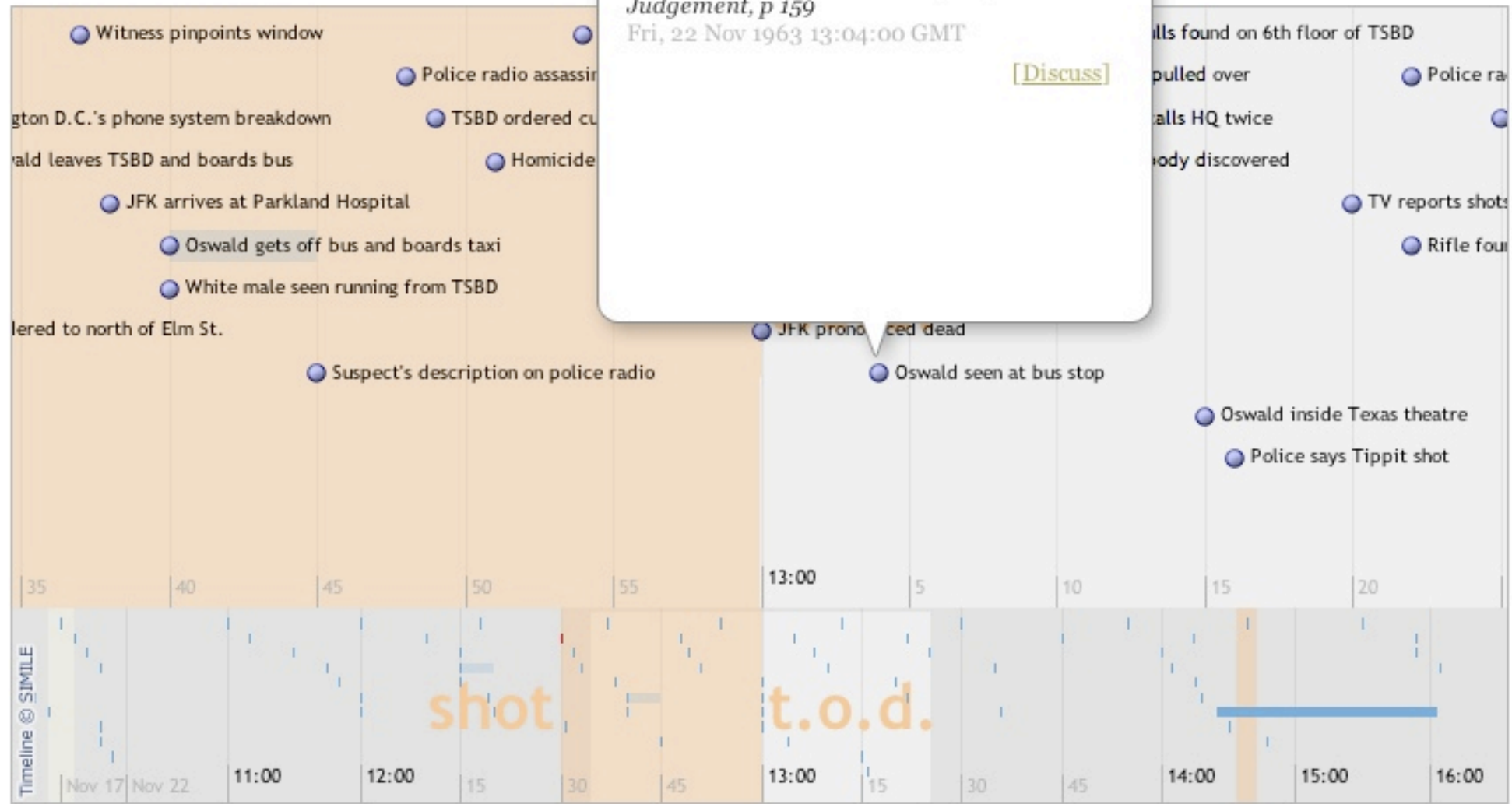
Timeline

Timeline is a DHTML-based AJAXy widget for visualizing t
example that you can play with. Pan the timeline by draggin

Oswald seen at bus stop

Oswald seen standing at bus stop outside his rooming house by his landlady. *ref. Rush to Judgement, p 159*
Fri, 22 Nov 1963 13:04:00 GMT

[\[Discuss\]](#)

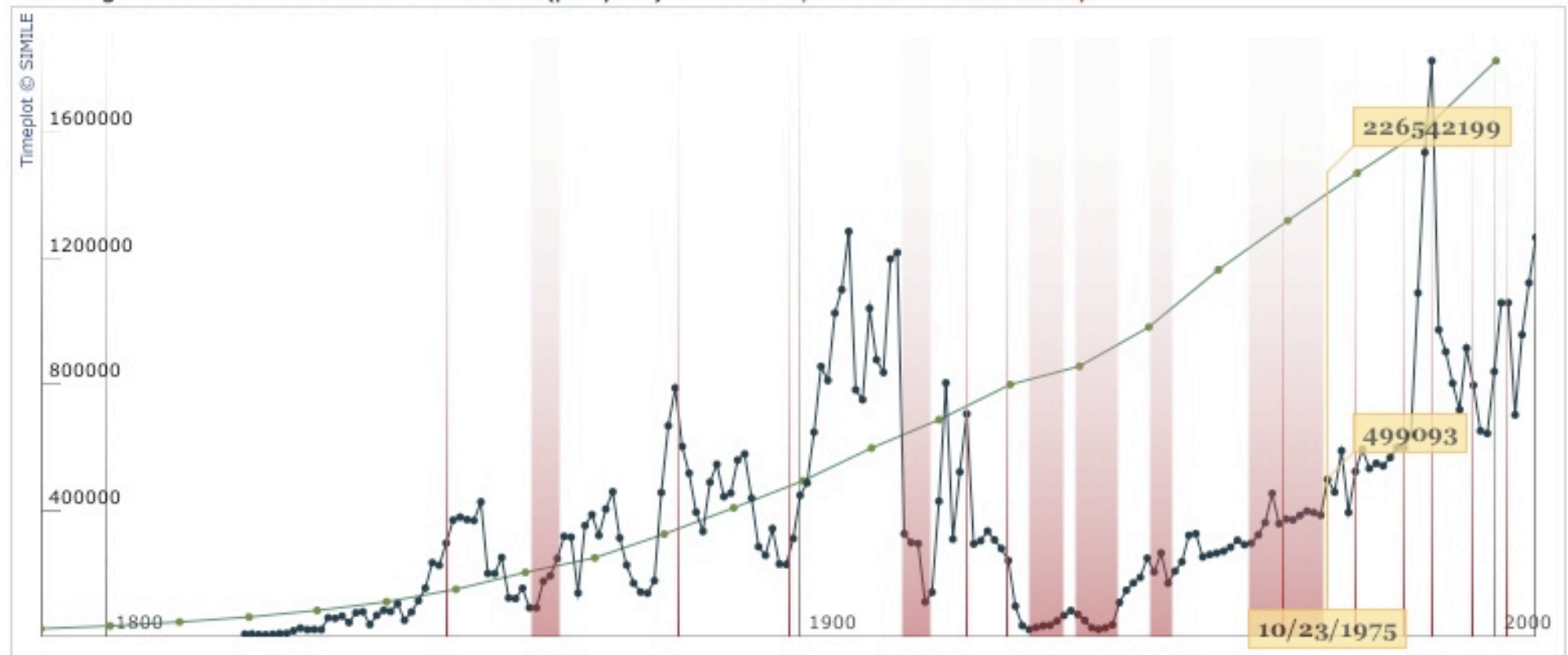


Timeplot

Timeplot is a DHTML-based AJAXy widget for plotting time series and overlay time-based events over them (with the same data formats that [Timeline](#) supports).

Here is a live example:

New Legal Permanent Residents in the U.S. (per year) vs. U.S. Population vs. U.S. History



Sources: [U.S. Department of Homeland Security](#), [U.S. Census Bureau](#) and [Wikipedia](#)

TOPHER'S BREAKFAST CEREAL CHARACTER GUIDE

Please refer to [Topher's original site](#) for copyright information. We are grateful to Topher for letting us host this data on our site.

Here is the [Exhibit JSON data file](#).

Search

49 Characters

Brands

- 3 General Mills
- 9 Kellogg's
- 2 Nabisco
- 4 Nestle
- 9 Post
- 12 Quaker Oats
- 10 Ralston

General Mills (3)



Trix Rabbit



Count Dracula



BuzzBee

sorted by: [brand](#); [then by...](#) • ☒ grouped as sorted

THUMBNAILS • TIMELINE

Countries

- 1 Canada
- 3 France
- 2 Great Britain
- 1 Japan
- 1 Spain
- 41 USA

Decades

- 1 1930
- 1 1940
- 4 1950
- 10 1960
- 4 1970
- 8 1980
- 8 1990
- 3 2000
- 10 unknown

Kellogg's (9)



Snap!



Tony the Tiger



Cornelius



Baby Toucan

Forms

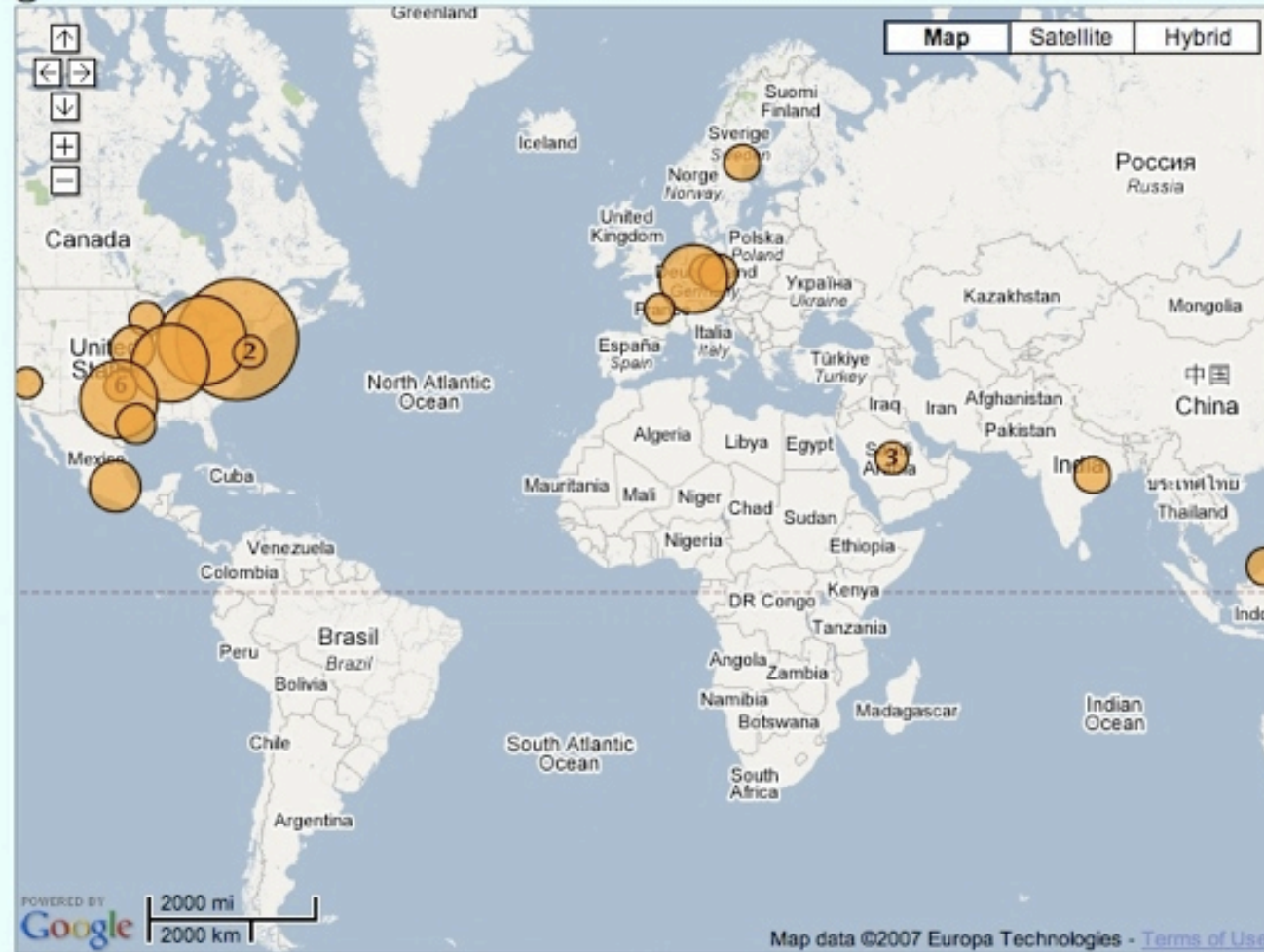
- 3 bear
- 3 beast
- 1 bee
- 5 boy
- 5 cartoon
- 1 cow
- 1 dinosaur
- 1 dog
- 1 frog
- 1 ghost
- 1 hippo
- 1 lego
- 1 monkey



Billionaires in History

WHERE THEY ARE FROM • COMPARE THEIR WEALTH

30 Billionaires



Search:

Companies

- 1 Aldi
- 1 American Fur Company
- 1 Berkshire Hathaway
- 1 Carnegie Steel Company
- 1 Ford Motor Company
- 1 Getty Oil Company
- 1 Hunt Refining Company
- 1 IKEA
- 1 Intel Corporation

Countries of Origin

- 1 Brunei
- 1 France
- 3 Germany
- 1 India
- 1 Mexico
- 3 Saudi Arabia
- 1 Scotland
- 1 Sweden
- 18 United States

about

contact

team

community

solutions

3D

publications

70 Items total.

WHAT • WHERE • WHEN

sorted by: year; then by... • grouped as sorted

1. **Eric Miller on Practical Applications of the Semantic Web**
Michelle Manafy in *EContent* , April 2007
2. **Brian Sletten publishes REST chapter in NoFluffJustStuff Anthology Vol. II**
Brian Sletten in *No Fluff Just Stuff Anthology Vol. II* , April 2007
3. **Eric Miller talks with Talis about the Semantic Web and Zepheira**
Paul Miller in *Talis* , April 2007
4. **The Future of the Web - CEO Guide To Technology**
Rachael King in *BusinessWeek* , April 2007
5. **Q&A with Tim Berners-Lee - CEO Guide To Technology**
Rachael King in *BusinessWeek* , April 2007
6. **Taming the World Wide Web - CEO Guide To Technology**
Rachael King in *BusinessWeek* , April 2007
7. **Weaving a Web Around Information - CEO Guide To Technology**
in *BusinessWeek* , April 2007
8. **The Semantic Web Podcast - CEO Guide To Technology**

Type

1 ✓ Book
4 ✓ News Article
3 ✓ Podcast
9 ✓ Presentation
51 ✓ Publication
1 ✓ Slideshow

Author

1 ✓ Brian Sletten
17 ✓ Eric Miller
1 ✓ Eric Neumann
1 ✓ Frank Manola
1 ✓ James Hendler
1 ✓ John Rodland

Audience

6 ✓ Business
1 ✓ Energy
60 ✓ General
2 ✓ Library Science
3 ✓ Life Sciences
1 ✓ Software Developers

about

contact

team

community

solutions

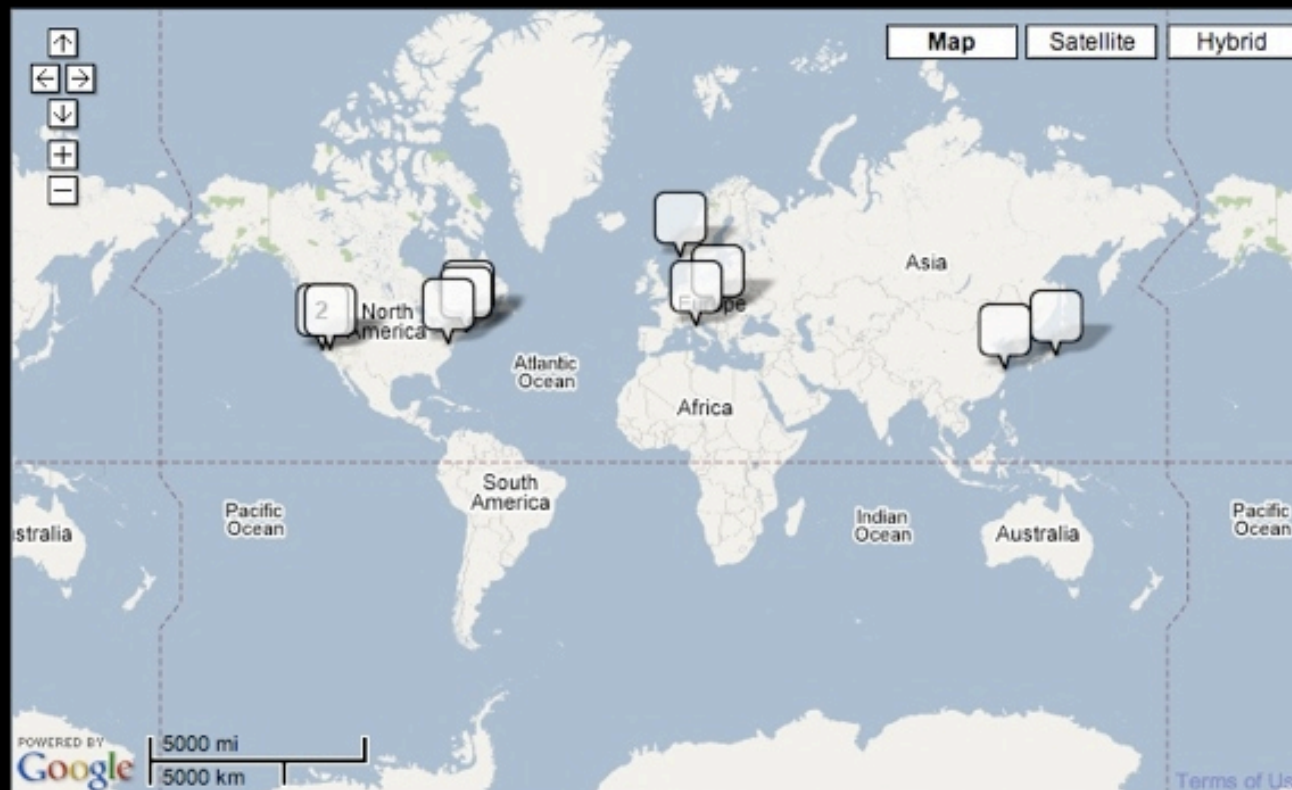
3D

publications

WHAT • WHERE • WHEN

70 Items total.

60 results out of 70 cannot be plotted.



Type

- 1✓ Book
- 4✓ News Article
- 3✓ Podcast
- 9✓ Presentation
- 51✓ Publication
- 1✓ Slideshow

Author

- 1✓ Brian Sletten
- 17✓ Eric Miller
- 1✓ Eric Neumann
- 1✓ Frank Manola
- 1✓ James Hendler
- 1✓ John Borland

Audience

- 6✓ Business
- 1✓ Energy
- 60✓ General
- 2✓ Library Science
- 3✓ Life Sciences
- 1✓ Software Developers

about

contact

team

community

solutions

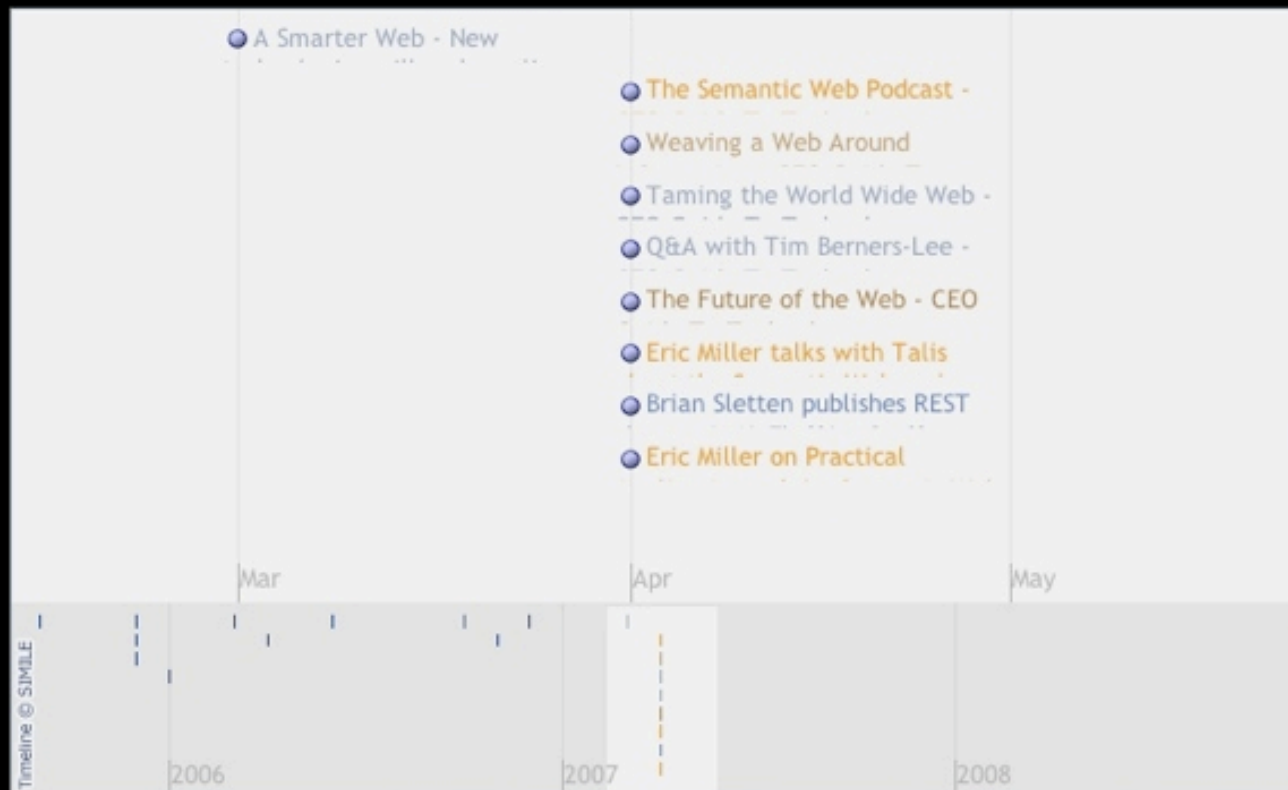
3D

publications

WHAT • WHERE • WHEN

70 items total.

2 results out of 70 cannot be plotted.



Type

- 1✓ Book
- 4✓ News Article
- 3✓ Podcast
- 9✓ Presentation
- 51✓ Publication
- 1✓ Slideshow

Author

- 1✓ Brian Sletten
- 17✓ Eric Miller
- 1✓ Eric Neumann
- 1✓ Frank Manola
- 1✓ James Hendler
- 1✓ John Berland

Audience

- 6✓ Business
- 1✓ Energy
- 60✓ General
- 2✓ Library Science
- 3✓ Life Sciences
- 1✓ Software Developers

back to business

- accelerate collaboration and knowledge exchange
- minimize data duplication / enable data reuse
- connect, reuse, republish
- move effective Business Intelligence
- adaptable systems to changing business needs



business as a web

- the Web's problems are Enterprise problems writ large
- silos, change, multiple participants, formats, too much data, etc.
- brittle systems
- “solved” by common addressing, linking, data frameworks



about

contact

team

community

solutions

3D

publications

70 Items total.

WHAT • WHERE • WHEN

sorted by: year; then by... • grouped as sorted

1. **Eric Miller on Practical Applications of the Semantic Web**
Michelle Manafy in *EContent* , April 2007
2. **Brian Sletten publishes REST chapter in NoFluffJustStuff Anthology Vol. II**
Brian Sletten in *No Fluff Just Stuff Anthology Vol. II* , April 2007
3. **Eric Miller talks with Talis about the Semantic Web and Zephaira**
Paul Miller in *Talis* , April 2007
4. **The Future of the Web - CEO Guide To Technology**
Rachael King in *BusinessWeek* , April 2007
5. **Q&A with Tim Berners-Lee - CEO Guide To Technology**
Rachael King in *BusinessWeek* , April 2007
6. **Taming the World Wide Web - CEO Guide To Technology**
Rachael King in *BusinessWeek* , April 2007
7. **Weaving a Web Around Information - CEO Guide To Technology**
in *BusinessWeek* , April 2007
8. **The Semantic Web Podcast - CEO Guide To Technology**

Type

- 1 ✓ Book
- 4 ✓ News Article
- 3 ✓ Podcast
- 9 ✓ Presentation
- 51 ✓ Publication
- 1 ✓ Slideshow

Author

- 1 ✓ Brian Sletten
- 17 ✓ Eric Miller
- 1 ✓ Eric Neumann
- 1 ✓ Frank Manola
- 1 ✓ James Hendler
- 1 ✓ John Berland

Audience

- 6 ✓ Business
- 1 ✓ Energy
- 60 ✓ General
- 2 ✓ Library Science
- 3 ✓ Life Sciences
- 1 ✓ Software Developers



- [About me](#)
- [Education, positions](#)
- [Publications](#)
- [Major Presentations](#)
- [R&D Activities](#)
- Professional activities:
 - [Current](#)
 - [Past](#)
- [Computing experiences](#)

Search facets:

year

- 12 ✓ 2004
- 12 ✓ 2005
- 13 ✓ 2006
- 9 ✓ 2007

country

Public Presentations (Since 2004)

(An [RDF](#) version of the list are also available.)

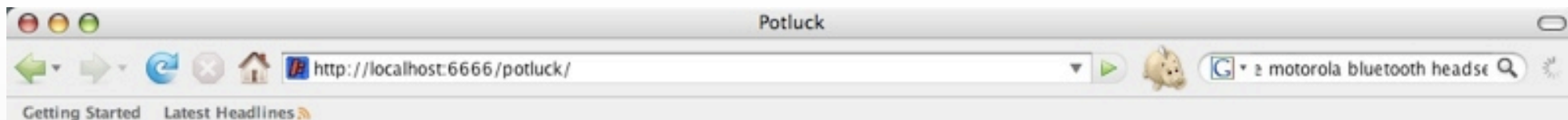
[DETAILS](#) • [CITIES ON A MAP](#)

46 presentation total

[Copy All](#)

sorted by: [date](#); [then by...](#) • ☐ grouped as sorted • ☐ show duplicates

1. "State of the Semantic Web", at "[Semantic Days 2007](#)", Stavanger, Norway, 2007-04-24. Slides are [available on-line](#).
2. "Introduction to the Semantic Web", at "[Semantic Days 2007](#)", Stavanger, Norway, 2007-04-23. Slides are [available on-line](#).
3. "Semantic Web: Anspruch und Wirklichkeit", at "[Handlungsschemata als Grundlage visueller und begrifflicher Strukturierung in der Wissensrepräsentation](#)" (Visual and Conceptual Structuring of Action-dependent Knowledge Representation), Paderborn, Germany, 2007-04-19. Slides are [available on-line](#).
4. "State of the Semantic Web", organized by the Tampere University of Technology and the W3C Finnish Office, Tampere, Finland, 2007-04-03. Slides are [available on-line](#).
5. "The Vision of the Semantic Web", at "[Infotech for Pharma & Biotech Europe 2007](#)", London, UK, 2007-03-14. Slides are [available on-line](#).
6. "Semantic Web", at "[ZKI Frühjahrstreffen](#)", Dortmund, Germany, 2007-03-06. Slides are [available on-line](#).
7. "State of the Semantic Web", at "[International Conference on Semantic Web & Digital Libraries](#)", Bangalore, India, 2007-02-23. Slides are [available on-line](#).
8. "Introduction to the Semantic Web", at "[International Conference on Semantic Web & Digital Libraries](#)", Bangalore, India, 2007-02-21. Slides are [available on-line](#).
9. "The offer and promises of Semantic Web", at "[European W3C Symposium on eGovernment](#)", Gijón, Spain, 2007-02-02. Slides are [available on-line](#).
10. "Semantic Web @ W3C: Activities, Recommendations and State of Adoption", at "[Industrial Track of the 5th International Semantic Web Conference \(ISWC2006\)](#)", Athens, GA, USA, 2006-11-09. Slides are [available on-line](#).



Potluck

a tool for mixing data

Paste the addresses of [Exhibit](#)-powered pages containing data to mix, one address per line. Then click Mix Data.

```
http://zepheira.com/publications/  
http://homepages.cwi.nl/~ivan/AboutMe/CV/Pres.html
```

Mix Data

Don't mix large exhibits. Potluck only handles a few hundred items.

For a working example, copy and paste these two URLs below into the text box above and click Mix Data:

```
http://simile.mit.edu/exhibit/examples/CSAIL-PIs/CSAIL-PIs.html  
http://ccnmtl.columbia.edu/draft/maurice/exhibit/ccnmtl_staff.html
```

For testing the Map view, try this URL:

```
http://cscie12.dce.harvard.edu/lecture_notes/2006-07/20070320/senate/senate.html
```

If you want to paste in a URL to another type of data, put its mime type after its URL separated by a space, e.g.,

```
http://people.csail.mit.edu/dhuynh/projects/test-data/presidents.xls xls
```

Supported mime types include: `rdf/xml`, `n3`, `xls`, `bibtex`.



[Not sure what's going on? Watch this screencast first \(65Mb\).](#)

Find: pairing

Next

Previous

Highlight all

Match case

Done

Potluck

Data origins: Zepheira :: Publications web site and Ivan Herman, Publications web site

VIEW AS: **TABLE** • MAP

☐ condense details • ☒ expand all details • ☐ collapse all details

Facets

New Column

drag
and
drop
fields
to
create
columns

Details

1.

A Smarter Web

label : A Smarter Web
origin : Zepheira :: Publications web site
audience : General
author : Mark Frauenfelder
date : 2001-11
identifier : http://www.technolog ... rticle.aspx?id=12640
month : Nov
publisher : The MIT Technical Review
title : A Smarter Web
year : 2001
type : News Article

2.

The offer and promises of Semantic Web

label : The offer and promises of Semantic Web
origin : Ivan Herman, Publications web site
LatLong : 43.5,-5.7
city : Gijón
country : Spain
date : 2007-02-02
event : European W3C Symposium on eGovernment
eventURI : http://www.w3c.es/Eventos/2007/eGov/
id : X_2007-02-02-597
slides : http://www.w3.org/2007/Talks/0202-Gijon-IH/
talkType : talk
title : The offer and promises of Semantic Web
year : 2007
type : presentation

origin

46 ✓ Ivan Herman, Publications web site
70 ✓ Zepheira :: Publications web site

drag and drop fields
to create facets

Data origins: Zepheira :: Publications web site and Ivan Herman, Publications web site

Data origins: Zepheira :: Publications web site and Ivan Herman, Publications web site

VIEW AS: [TABLE](#) • [MAP](#)

Satellite

Satellite

Hybrid

origin

46 ✓ Ivan Herman,
Publications web site

70 ✓ Zepheira :: Publications
web site

drag and drop **fields** to
create facets

Address

drop fields here

Latitude/Longitude

drop fields here

Color Coding

drop fields here

Potluck

Data origins: Zepheira :: Publications web site and Ivan Herman, Publications web site

You've just done: Add field locLatLng as latitude/longitude to map by [Undo](#)

VIEW AS: TABLE • **MAP**



Facets

origin

- 46 ✓ Ivan Herman, Publications web site
- 70 ✓ Zepheira :: Publications web site

drag and drop fields to create facets

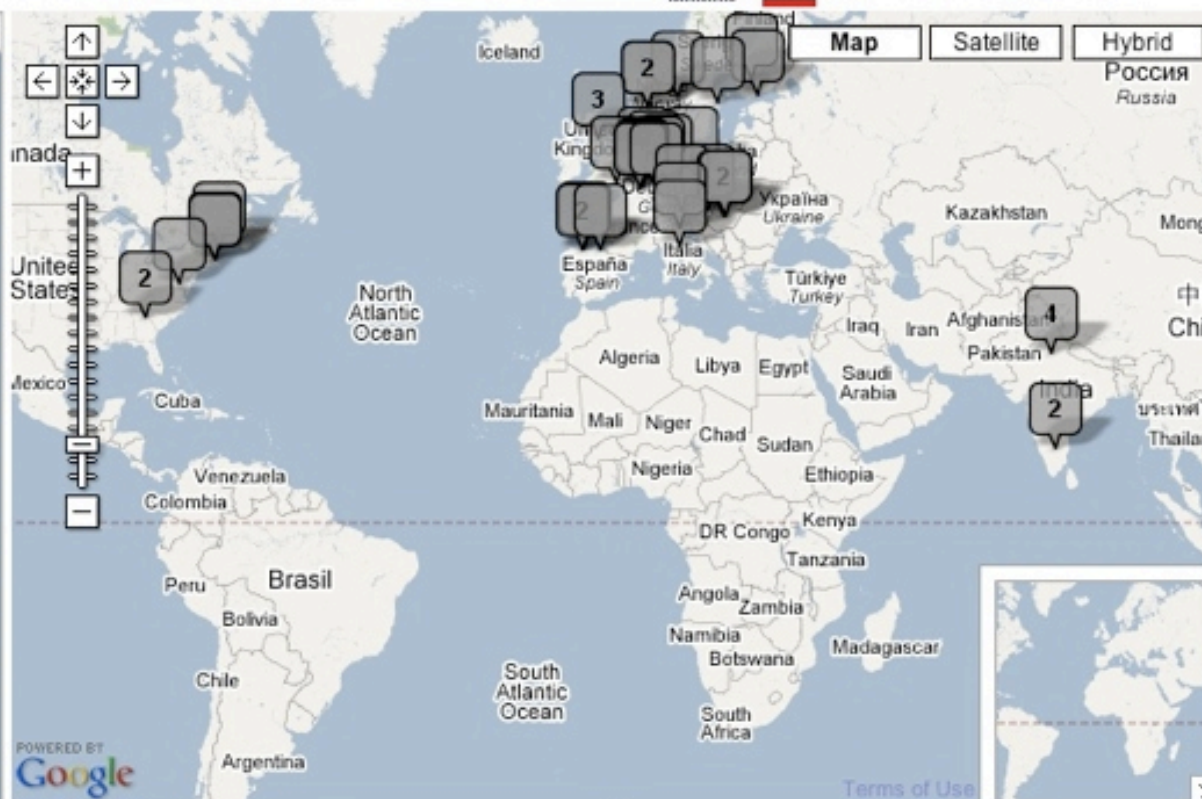
Potluck

Data origins: Zepheira :: Publications web site and Ivan Herman, Publications web site

You've just done: Add field LatLong as latitude/longitude to map by [Undo](#)

VIEW AS: [TABLE](#) • **MAP**

- label
- origin
- type
- URI
- city
- country
- date
- event
- eventEnglish
- eventURI
- id
- LatLong
- organizer
- slides
- talkType
- title
- year
- audience
- author
- conference
- date
- id
- identifier
- issue
- journal
- location
- locLatLng
- month
- pages
- pretype



Facets

origin

- 46 ✓ Ivan Herman, Publications web site
- 70 ✓ Zepheira :: Publications web site

drag and drop fields to create facets

Missing Mixed

Address
drop fields here

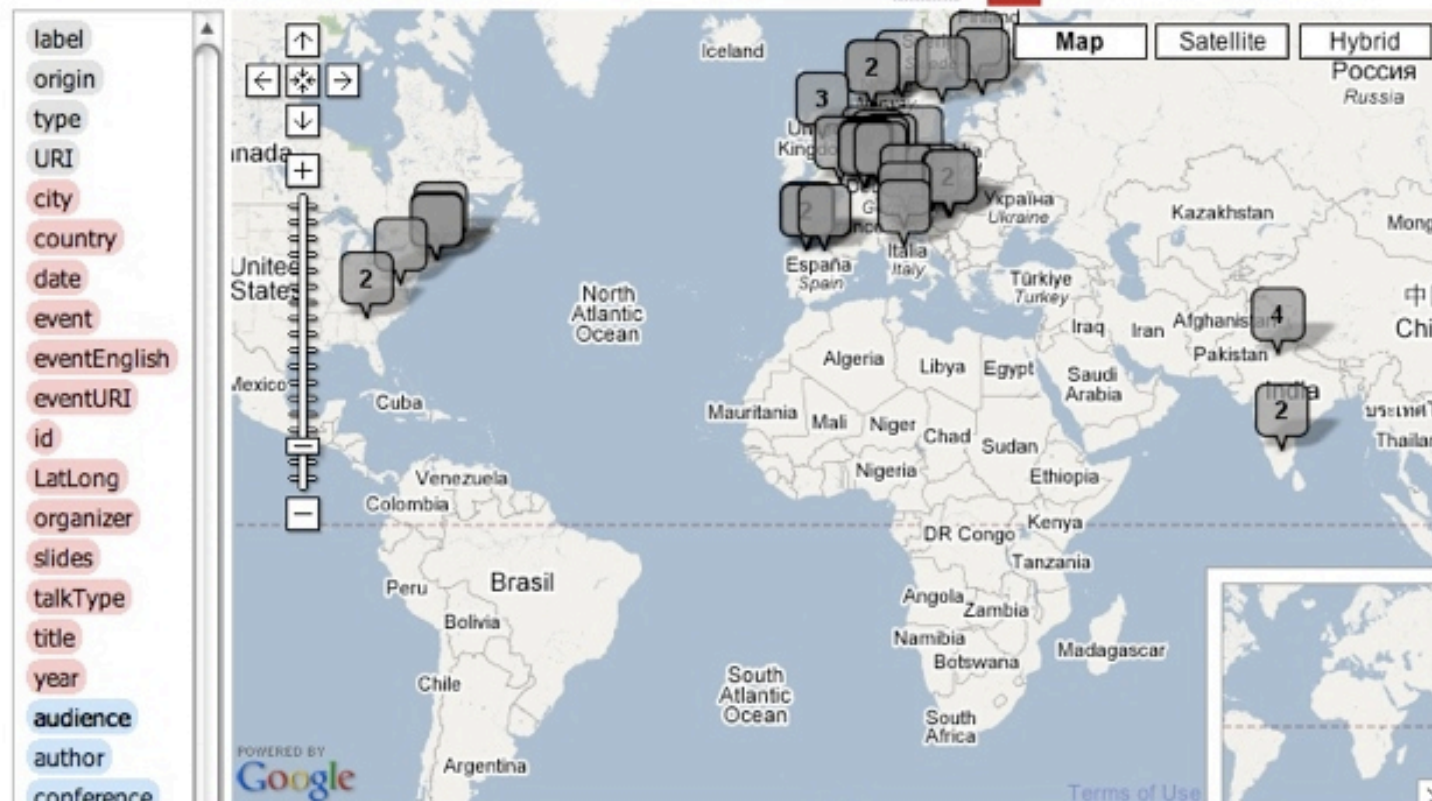
Latitude/Longitude
locLatLng ×
LatLong ×
drop fields here

Color Coding
drop fields here

Data origins: Zepheira :: Publications web site and Ivan Herman, Publications web site

Data origins: Zepheira :: Publications web site and Ivan Herman, Publications web site

You've just done: Merge talkType into facet pretype / type [Undo](#)

VIEW AS: [TABLE](#) • [MAP](#)

Facets

origin

46 ✓ Ivan Herman,
Publications web site

70 ✓ Zepheira :: Publications
web site



```
pretype  type  talkType
```

- 4 ✓ invited talk
- 7 ✓ Keynote
- 3 ✓ keynote
- 1 ✓ s0-Book
- 4 ✓ s0-News Article
- 3 ✓ s0-Podcast
- 9 ✓ s0-Presentation

drag and drop **fields** to
create facets

Address
drop fields here

Latitude/Longitude

locLatLng LatLong 

drop fields here

Color Coding
drop fields here

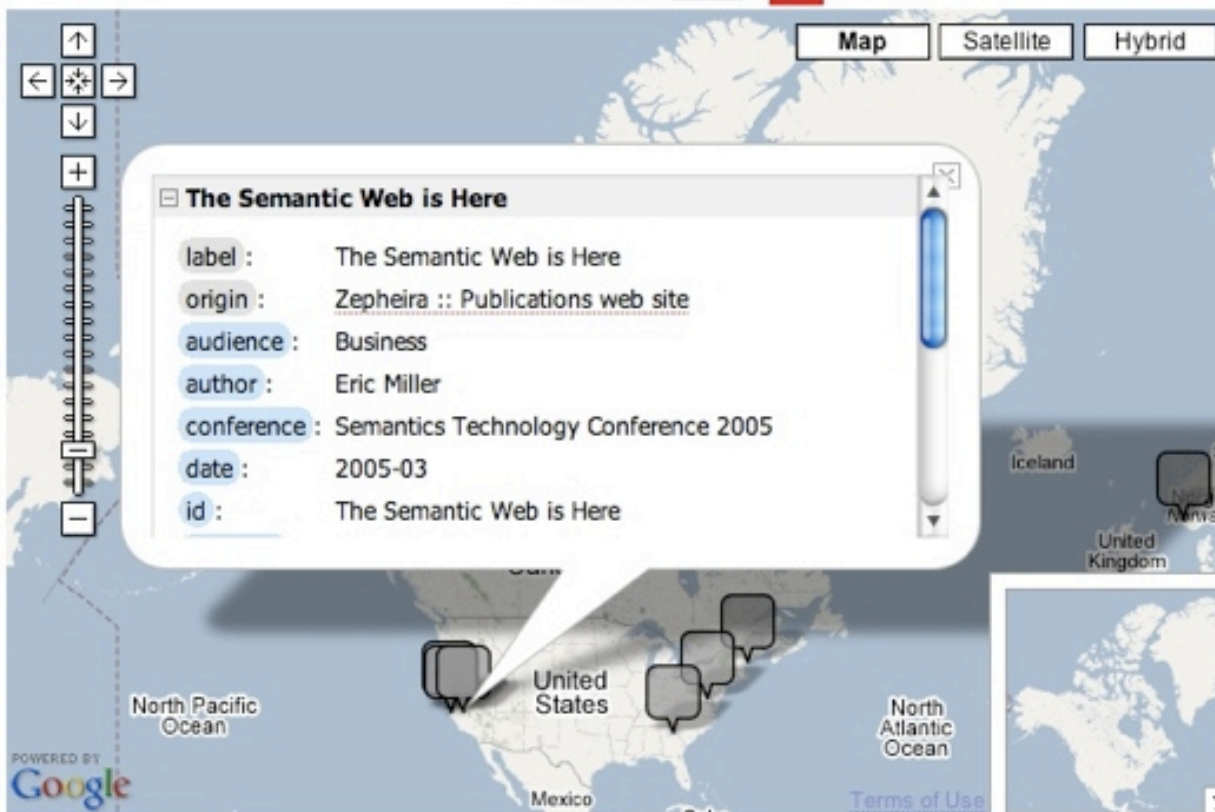
Potluck

Data origins: Zepheira :: Publications web site and Ivan Herman, Publications web site

You've just done: select keynote Undo

VIEW AS: TABLE • MAP

label
origin
type
URI
city
country
date
event
eventEnglish
eventURI
id
LatLong
organizer
slides
talkType
title
year
audience
author
conference
date
id
identifier
issue
journal
location
locLatLng
month
pages
pretype



Facets

origin

- 3 ✓ Ivan Herman, Publications web site
- 7 ✓ Zepheira :: Publications web site

pretype type talkType 2 ✓

- 4 invited talk
- 7 ✓ Keynote
- 3 ✓ keynote
- 1 s0-Book
- 4 s0-News Article
- 3 s0-Podcast
- 9 s0-Presentation

drag and drop fields to create facets

Address
drop fields here

Latitude/Longitude
locLatLng
LatLong
drop fields here

Color Coding
drop fields here

empowering users

- share data
- stitch data together “as needed”
- share linkages
- make more effective decisions



empowering enterprises

- share data
- stitch data together “as needed”
- share linkages
- make more effective decisions



accelerating communities

- collective intelligence



accelerating business

- coordinated intelligence



which translates into..

- more effective business intelligence
- more effective discovery
- less redundancy / duplication
- agility, adaptability
- decreased costs / increased savings



a perfect('esq) storm

- business and IT strategies can no longer afford to be misaligned
- traditional corporate IT strategies are not as effective in dynamic, changing environments
- costs for leveraging Web technologies for managing data are less than the cost not to



the web

- strategic development platform
- combination of enabling (open!) standards and technologies
 - greater than the sum of individual parts
- web evolves as communities evolve
- shaping each other



breaking boundaries

- communities
- corporations
- governments



creating

- new business models and opportunities
- new ways of sharing
- new ways of learning



and that is a very
good thing :)



obstacles?



several...
but here are a few



challenge :: education



good news ...



growing awareness in
business, media,
conferences, etc.



your listening



i hope :)



challenge :: content



data dilemma

- costs for creating non-interoperable data radically dropping
- need for integrating data radically increasing



content creation

- example: wiki, blogs, etc.
- small footprint (easy to install, create)
- easy to miss (from IT standpoint)
- easy to create many stove-pipes (from KM standpoint)
- regulatory nightmare



Same Data, Different Form



**Human
Readable**

**Machine
Readable**



many options

- microformats
- GRDDL
- RDF/A



good news ...







enterprise web of data

- wiki, blog, templates
- wrappers - expose existing data stores
- ground-up, as needed data integration
- flexible architectures to adapt to changing business needs (and indicators where new ones can be found)



challenge :: persistence



good news ...



purlz.org



challenge :: governance



who to trust?



data policies



good news ...



same data framework



challenge :: interfaces



how to take advantage of
recombinant data



but hide complexity to
end user



good news ...



simile.mit.edu



many others...



we now have many of the
design pieces in place



and tools are being built
even as we speak



can you afford to ignore
it?



more important -
imagine the potential



thank you



questions?

Eric Miller

em@zepheira.com



appendix



data dilemma

- costs for creating non-interoperable data radically dropping
- need for integrating data radially increasing



content creation

- wiki, blogs
- simple footprint
- easy to install, easy to create
- from IT management standpoint, easy to miss



wiki, blogs

- inward facing
- outward facing
- regulatory issues
- painfully able to create more silos



microformats

“Designed for humans first and machines second, microformats are a set of simple, open data formats built upon existing and widely adopted standards. Instead of throwing away what works today, microformats intend to solve simpler problems first by adapting to current behaviors and usage patterns”

—microformats.org



Same Data, Different Form



**Human
Readable**

**Machine
Readable**



basic idea

- embed minor, specialized variations within established vocabularies
- start with base schema such as HTML
- define specialized variations
 - conventions for attribution, friends, calendars, etc.



example

Google nofollow

<p>Nice blog. Buy your medz here</p>

XFN

<div class='blogroll'>
 Brian
</div>

vote-links

<p>We decided not to implement XQuery...</p>



a seriously mixed bag

- the good: rough consensus is good, even for minor syntactic details
- the bad: it's very easy for microformats to clash, because of lack of Semantic grounding
- the ugly: formats get ugly real fast (e.g. XOXO) when you try to tunnel one set of semantic into another



GRDDL

- “Gleaning Resource Descriptions from Dialects of Languages”
- W3C Proposed Recommendation for using transforms such as XSLT to convert XML to RDF
- piggy-backs off microformat profiles



example

Main GRDDL declaration

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head profile="http://www.w3.org/2003/g/data-view">
    <title>Some Document</title>
    <link rel="transformation"
      href="http://www.w3.org/2000/06/dc-extract/dc-extract.xsl" />
    <meta name="DC.Subject" content="Semantic Search" />
    ...
  </head>
  ...
</html>
```

link to GRDDL transform



result

```
<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <rdf:Description rdf:about="">
    <dc:title>Some Document</dc:title>
    <dc:subject>Semantic Search</dc:subject>
  </rdf:Description>
</rdf:RDF>
```

Gleaned from XHTML



RDF/A

- RDF/A is a system for embedding data right into XHTML documents using similar ideas to microformats
- Eliminates extra transform step required by GRDDL to extract RDF



microformat example

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Some Document</title>
  </head>
  <body>
    ...
    <p>This document is licensed under a
      <a rel="license" href="http://creativecommons.org/licenses/by-nc/2.5/">
        Creative Commons Non-Commercial License </a>.
    </p>
    ...
  </body>
</html>
```



RDF/A example

```
<html xmlns="http://www.w3.org/1999/xhtml"  
  xmlns:cc="http://creativecommons.org/licenses/">  
  <head>  
    <title>Some Document</title>  
  </head>  
  <body>  
    ...  
    <p>This document is licensed under a  
      <a rel="cc:license" href="http://creativecommons.org/licenses/by-nc/2.5/">  
        Creative Commons Non-Commercial License </a>.  
    </p>  
    ...  
  </body>  
</html>
```

